PolicyIE Annotation Guideline

In order to facilitate the research of Natural Language Processing (NLP) and help the advance of AI reading comprehension in privacy policy, we ask you to perform annotations. This guideline describes how you are going to annotate the privacy policy corpus and provide examples and rules of the annotations.

Annotation Schema

In this section we describe your annotations. The task is to identify privacy practices (intents) and the corresponding slots after reading each sentence of the online privacy policy document. Intent classification and slot filling in our corpus is the task of interpreting the privacy practices by extracting the intents and the relevant slots. In our annotations, We consider the following five intents and the relevant slots.

1. **Data Collection/Usage**
    a. Slot 1 Action: action that indicates the data are collected;
    b. Slot 2 Data Collector: entities who collect/use the data on behalf of service providers;
    c. Slot 3 Data Provider: entities who provide the data;
    d. Slot 4 Data collected: data that are being collected/used;
    e. Slot 5 Polarity: whether the privacy practice is enforced;
    f. Slot 6 Condition: how/under some conditions the data are collected (e.g., when you sign the agreement);
    g. Slot 7 Purpose: reasons why the data are being collected/used.

2. **Data Sharing/Disclosure**
    a. Slot 1 Action: action that indicates the data are shared;
    b. Slot 2 Data Sharer: entities who share the data on behalf of service providers to third party entities;
    c. Slot 3 Data Provider: entities who provide the data;
    d. Slot 4 Data Receiver: third party entities who receive the data;
    e. Slot 5 Data Shared: data that are being shared/disclosed;
    f. Slot 6 Polarity: whether the privacy practice is enforced;
    g. Slot 7 Condition: how/under some conditions the data are shared;
    h. Slot 8 Purpose: reason why the data are being shared/disclosed.

3. **Data Retention/Storage**
    a. Slot 1 Action: action that indicates the data are retained/stored;
    b. Slot 2 Data Holder: entities who retain/store the data on behalf of service providers;

c. Slot 3 Data Provider: entities who provide the data;
d. Slot 4 Retention Period: how long the data are being retained;
e. Slot 5 Storage Place: where the data are stored;
f. Slot 6 Data Retained: data that are being shared/disclosed;
g. Slot 7 Polarity: whether the privacy practice is enforced;
h. Slot 8 Condition: how/under some conditions the data are retained/stored;
i. Slot 9 Purpose: reason why the data are being shared/disclosed.

4. Data Security/Protection

a. Slot 1 Action: action that indicates the data are secured/protected;
b. Slot 2 Data Protector: entities who secure/protect the data on behalf of service providers;
c. Slot 3 Data Provider: entities who provide the data;
d. Slot 4 Data protected: data that are being shared/disclosed;
e. Slot 5 Protect Against: security threats that the practice is against;
f. Slot 6 Protection method: data protection method (e.g., user authentication, etc.);
g. Slot 7 Polarity: whether the privacy practice is enforced;
h. Slot 8 Condition: how/under some conditions the data are secured/protected;
i. Slot 9 Purpose: reason why the data are being secured/protected.

5. Other

Except “Other”, each intents consist of a list of slots. Note that a sentence might contain only one/more slots and not every slot appears in each annotation. We categories the slots into two types: Type-I slots are Type-II slots. Type-II slots include purposes, conditions, polarity and Protection method and other slots belong to Type-I slots. Slots might also have attributes. The following table shows the slots and the corresponding attributes.

Slots	Attributes
Action	None
Data Provider	(1) User (2) Third party entity
Data Collector	(1) First party entity
Data Collected	(1) General Data (2) Aggregated/Non-identifiable data (3) Contact data (4) Financial data (5) Location data (6) Demographic data (7) Cookies, web beacons and other technologies (8) Computer/Device data (9) User online activities/profiles (10) Other data
Data Sharer	(1) First party entity
Data Shared	(1) General Data (2) Aggregated/Non-identifiable data (3) Contact data (4) Financial data (5) Location data (6) Demographic data (7) Cookies, web beacons and other technologies (8) Computer/Device data (9) User online activities/profiles (10) Other data

Data Receiver	(1) Third party entity
Data Holder	(1) First party entity (2) Third party entity
Data Retained	(1) General Data (2) Aggregated/Non-identifiable data (3) Contact data (4) Financial data (5) Location data (6) Demographic data (7) Cookies, web beacons and other technologies (8) Computer/Device data (9) User online activities/profiles (10) Other data
Storage Place	None
Retention Period	None
Data Protector	(1) First party entity (2) Third party entity
Data Protected	(1) General Data (2) Aggregated/Non-identifiable data (3) Contact data (4) Financial data (5) Location data (6) Demographic data (7) Cookies, web beacons and other technologies (8) Computer/Device data (9) User online activities/profiles (10) Other data
Protect Against	Security threat
Purpose	(1) Basic service/feature (2) Advertising/Marketing (3) Legal requirement (4) Service operation and security (5) Personalization/customization (6) Analytics/research (7) Communications (8) Merge/Acquisition (9) Other purpose
Condition	None
Polarity	(1) Negation
Protection Method	(1) General safeguard method (2) User authentication (3) Access limitation (5) Encryptions (6) Other protection method

Attribute Descriptions and Examples

Next we provide description and examples (including some difficult corner cases) to the slot attributes in detail. Please go to our annotation url for more annotation examples.

First-party-entity: The first-party company or organization that collects, uses, shares, protects and secure user's data (e.g., we, company name).

Third-party-entity: The third-party involved in data practices such as the data sharing with third parties or data collection by third parties. A third party is referred to as the company / organization other than the first party company/organization described in the privacy policy.

User: The customer or client involved in the privacy policy (e.g., you, children under 13).

General Data: The information/data that is collected, shared, etc., but the type of information/data is not specified or mentioned (e.g., general information, personal information).

Non-identifiable/aggregated data: The information/data that is collected, shared, etc., and it is anonymized (e.g., connection with the user's identity is removed) or aggregated (e.g, combined with user's information so that it is not possible to uniquely identify the identity of the user) (e.g., aggregated data, non-identifiable information).

Identification/Contact Information: Identifiers that uniquely identify a person or data that helps contact the user (e.g., SSN, driving license number, name, email address, phone number, street address).

Financial Information: user's financial information (e.g., credit/debit card data, payment information, credit scores).

Location Information: user's geo-location information (e.g., user's current location regardless of granularity, i.e., could be the exact location, ZIP code, city-level).

Demographic Information: user's demographic Information (e.g., gender, age, occupation, education, education, etc.).

Cookies/Web beacons and other technologies: Identifiers locally stored on user's device by company/organization or third-parties including cookies, beacons, or similar that are commonly used to uniquely identify users, but that are not essential to establish a connection with the user's device or to provide a service.

Computer/Device Information: User's computer or device information (e.g., device IDs, MAC address, IP address,identifiers of the user's computer and mobile device, the type of operating system (OS) or web browser that the user uses, or similar computer or device information).

User Online Activities/Profile: The user's online activities on the first party website/app or other websites/apps (e.g., pages visited, time spent on pages, general user behavior online, etc.) or the user's profile on the first-party website/app and its contents (e.g., user profile, user comments, user profile preferences).

Other Data: Other types of information not covered by the above categories.

Basic Service/Feature: to provide basic service and features that the user explicitly requests and that is part of the website/app's basic service or functionality (e.g., payment processing, account registration).

Advertising/Marketing: to show ads that are either targeted to the specific user or not targeted or to contact the user to offer products, services, or other promotions (e.g., sending marketing emails, calling or texting users with marketing messages).

Legal Requirement: to comply with legal obligations, e.g., regulations, government data requests, government retention requests, law enforcement requests in general, etc.

Service Operation and Security: For website/app operation and security, enforcement of terms of service, fraud prevention, protecting users and property, etc.

Personalization/Customization: For providing users with a personalized experience (e.g., by allowing them to arrange how the website/app looks, based on the user's preferences or language, etc).

Analytics Research: For understanding the website/app's audience, improving the website/app, informing company strategy, or general research.

Communications: For communication and contact purposes (e.g., contacting users, making a complaint, resolving disputes, customer service activities).

Merge/Acquisition: If company/organization merges or is acquired it transfers users' information to another company/organization.

Other Purpose: Other specific purposes that are not covered above (e.g., the purpose described above, all other vague purposes, take other actions permitted by law or disclosed in this Privacy Policy).

General Safeguard Method: Security measures that the website/app implements to protect users' information, specific methods are not specified or mentioned (e.g., general steps, measures, safeguard).

User Authentication: As indicated by the name (e.g., password).

Access limitation: As indicated by the name.

Encryptions: As indicated by the name (e.g., SSL).

Other Protection Methods: Other specific methods that are not covered above.

Security Threat: Security threats that the security measures protect against (e.g., misuse, and unauthorized access, disclosure, alteration).

More Rules/Notes of Annotations

1. For some intents, there might not be any slot mentioned in the text.
2. An intent annotation must be associated with an action word. If you think multiple words could be a potential action, just tag the first word/first mention as the action. For example, in "*We collect and use …"*, tag "*collect*" (rather than use) as the action word.
3. We do not allow annotation in section header, which is the first line of the document. Note that in some sections, there might not be any section headers.
4. No space allowed in the beginning/ending of an annotation span.
5. Do not tag all mentions as slots. Do not annotate all mentions of an entity. For example, If there are multiple "*we*" mentioned in a sentence and "*we*" indicates the first party entity, we just tag the first mention of "*we*" as the first party entity.
6. Conventions to tag Type-I slots (e.g., data, first party entity, third party entity, user and different data types):
    a. Normal entities are usually be short phrases instead of long clauses;
    b. Whether we should include adjectives/modifiers for normal entities depends on whether the adjectives are able to largely change the semantic meaning of the entity. Please consider the following two examples for comparing the difference
        (1) some personal information
        (2) Children under 13
7. Conventions for tagging Type-II slots (e.g., condition and purpose):
    a. Do not consider conjunction (e.g., "*when*", "*if*") in the annotations;
    b. Purpose can take in the form of (but not limit to):
        i. *Noun/Noun phrases*: We may ask you to provide Personal Information for the purpose of understanding your personal needs better. (Purpose)
        ii. *Verb*: We may ask you to provide Personal Information to gain better knowledge of your needs. (Purpose)
    c. Condition can take in the form of (but not limit to):
        i. *Clauses*: *We may ask you to provide Personal Information when you use our websites.* (Condition)
        ii. Modifier: third parties who consent to our agreement (Condition)

2. In some cases, different intents can appear in the same sentence. Please see the following examples:

    "*We collect, use and share your information to …"*

    In the above example, the sentence contains two intents: *data collection/usage* and *data sharing/disclosure.*